图书情报工作
第64卷 第6期 2020年3月

全局视角下的网络社区多元知识关联挖掘*

肖璐[1] 赵之辉[1] 陈果[2]
[1] 南京财经大学新闻学院 南京 210023 [2] 南京理工大学经济管理学院 南京 210094

摘 要:[目的/意义]网络社区中存在多种知识单元,知识单元间又有错综复杂的关系。有必要在保留知识单元全局信息的前提下,统一、简洁地开展多元知识关联挖掘。[方法/过程]提出网络社区多元知识关联挖掘的实现方案。首先,将网络社区中3种典型知识单元(用户、文本、词语)及其在知识交流中多种关系抽取为超网络;其次,利用网络表示学习算法将超网络中节点表示为统一特征空间下的低维稠密向量;最后,基于节点的向量开展多元知识关联计算。[结果/结论]以丁香园心血管论坛为例开展实验,验证方案的有效性。该方案既保留知识单元的全部信息,知识关联的挖掘又在统一低维特征下开展,且最终所得的知识关联满足网络社区知识组织场景多样性的要求。

关键词:知识关联挖掘 超网络 网络表示学习 网络社区
分类号:G250
DOI:10.13266/j.issn.0252-3116.2020.06.012

1 引言

当前,网络社区用户与资源数量迅速积累,已成为用户知识交流和利用的重要场所。然而,网络社区普遍存在资源"碎片化"和知识组织"粗粒度化"的问题,导致大量有重要价值的知识淹没于海量数据中,难以被用户有效获取与利用。网络社区知识组织面临的一个关键问题是,知识单元[1]形式复杂且粒度差异明显,诸如用户、帖子、评论、主题等都是知识组织中需要涉及的知识单元,而同粒度与跨粒度知识单元之间又存在错综复杂的多种关系。有必要全局性、系统化地梳理这些关系,并在其基础上开展多元关联挖掘与揭示,以促进网络社区知识组织的深化、细化。

目前,网络社区知识关联发现以一元关系挖掘为主,多元关系挖掘主要利用超网络技术对多个一元关系进行多元描述,各类型关系相互独立,本质是一元关系的多元呈现。尚未真正实现全局视角下的多元关联挖掘,并且网络节点及关系的异构也导致跨粒度多元关系利用困难。

如何在保留全局视角的前提下,屏蔽网络社区跨粒度知识单元和异质关系的干扰,简洁、有效地揭示其中蕴含的多元知识关联?笔者提出一种方案:通过构造网络社区知识超网络以保障全局性,再利用网络表示学习将知识单元表示为形式一致的低维稠密向量,各种知识单元间的关联均可基于其向量计算获得。该方案可快速生成网络社区多元知识关联体系,用于指导网络社区知识组织,并以医学网络社区丁香园心血管论坛为对象,验证该方案的可行性与有效性。

2 相关研究

2.1 网络社区多粒度知识单元关系挖掘

网络社区主要包括用户、文本与词语3种粒度的知识单元,同粒度与跨粒度知识单元之间存在多种关系。其中一元关系挖掘包括:①用户-用户关系。细分为直接关系与间接关系。前者通过分析用户之间的关注、回复等行为得到[2],后者通过分析直接关联得到[3]。②文本-文本关系。利用文本相似度衡量,相似度计算中常用扩展源包括:领域本体[4]、搜索引擎[5-6]等。③词语-词语关系。标签是语词关系挖掘重要对象,关系强度主要从共现频次转化而来,学者利用社会网络分析[7]、LSA(latent semantic analysis)[8]等方法挖掘标签的语义层级关系,弥补语义缺失的不足;

* 本文系国家社会科学基金青年项目"学术型网络社区多元关联挖掘与知识聚合研究"(项目编号:16CTQ025)研究成果之一。
作者简介:肖璐(ORCID:0000-0001-5485-1407),讲师,博士,硕士生导师,E-mail:ahjk_xiaolu@163.com;赵之辉(ORCID:0000-0002-8192-4269),硕士研究生;陈果(ORCID:0000-0003-2873-1051),副教授,博士,硕士生导师。
收稿日期:2019-06-10 修回日期:2019-12-23 本文起止页码:100-107 本文责任编辑:徐健

除此之外,可利用主题识别或抽取技术从文本中自动识别特征或抽取新特征,并挖掘结构化语义关系[9]。④用户－文本关系。细分为直接关系与间接关系。前者根据用户对文本的发布、分享等行为信息转化得到,关系强度依据行为类型可人工给定阈值。后者通过文本－文本关系或用户－用户关系传递得到,实质是基于文本或用户的协同推荐[10]。⑤文本－词语关系。从文本信息中提取标签或特征词时,文本－词语关系就已建立。关系强度计算思路主要分为两种:一种是“0,1”,即只统计特征词在文本中出现与否;另一种利用特征权重(如IF-IDF、信息增益等)进行细粒度衡量。⑥用户－词语关系挖掘。通过用户－文本关系与文本－词语关系传递得到。

除此之外,学者尝试利用超网络技术进行多元关系融合。例如肖璐[11]构建了面向网络社区多粒度知识聚合的知识超网络,该网络包含词语、句子与文本3种粒度知识单元及共现、语法、包含、隶属等多元关系;王传清等[12]构建了用于数字资源深度聚合的数字资源超网络,该网络包含文献知识、著作权人、物质载体3种知识单元及引用、共现、耦合等多元关系。

2.2 网络表示学习在知识组织中的应用

网络表示学习在知识组织中的应用研究主要集中在3方面:①学者合作预测与论文影响力预测。张金柱等[13]利用网络表示学习从学者合著关系网络学习学者的特征向量,通过向量相似度计算进行学术合作预测;林原等[14]将表示学习综合应用在作者、关键词、机构、作者与关键词等多类共现网络中,在分析学者潜在合作可能性时,这种融合多元知识单元的方法克服了传统方法重点关注高产学者的不足;樊玮等[15]利用网络表示学习将论文、作者、期刊或会议3类知识单元映射到低维稠密向量空间,构建能较好还原网络局部结构信息的异构学术网络表示模型,更准确预测论文的影响力。②知识的表示学习。张潇鲲等[16]提出在文本信息网络表示学习中加入外部词向量,融合语义与结构特征进行文本的特征向量表示;朱国进等[17]构建了融合命名实体与词向量的网络文本表示学习模型;朱靖雯等[18]将网络表示学习应用于HowNet知识库中,实现跨语言与语义单位的向量表示。③社交网络用户关联分析。韩忠明等[19]利用网络表示学习对用户属性、网络结构等多类信息融合分析,得到用户特征向量,实现多角度用户关联挖掘;杨奕卓等[20]利用网络表示学习对用户名与拓扑结构信息进行融合分析,得到账号特征向量,实现跨网络用户身份匹配分析。

目前相关研究主要在现实需求驱动下,利用网络表示学习对多元关系进行融合挖掘,实现单一维度下的知识组织。已有研究成果较为零散,多元关联包含的知识单元与关系类型有限,难以支持全局视角下的知识多维组织。

3 基于网络表示学习的多元知识关联挖掘方案设计

3.1 基于网络表示学习的多元知识关联挖掘思路

多元知识关联体系是网络社区知识组织细化和知识服务深化的基础。目前,网络社区关系挖掘以一元关系为主,涉及多元关系挖掘的研究主要通过异构网络进行多元关联描述,尚未真正实现全局视角下的多元关联挖掘。网络表示学习是一种以初始网络为基础,将网络节点表征成具有推理能力的低维稠密向量的技术,节点的低维稠密向量表示在保留初始网络信息的同时实现了网络的重构[21]。因此,笔者提出的挖掘思路是:首先,利用超网络将多粒度知识单元及其多元关系描述在统一网络中,该网络是利用网络表示学习进行全局视角知识关联挖掘的基础。然后,利用网络表示学习技术将知识单元表征成结构统一的向量集合,一个向量代表一个知识单元,知识单元关联由向量相似度表征。笔者将这种知识单元及其关联集合称为网络社区多元关联体系,该关联体系能被计算机快速处理分析,可作为领域背景知识支持知识的多维组织。具体思路见图1。

3.2 基于网络表示学习的多元知识关联挖掘流程

网络社区多元知识关联体系是通过对网络社区知识超网络进行网络表示学习得到,而多粒度知识单元识别与多元关系挖掘又是知识超网络构建的基础,因此笔者提出的流程主要包括3部分,具体如图2所示。值得一提的是流程第一部分,将列举多粒度知识单元的多元关系挖掘方法,在实际应用中可根据网络社区特点确定具体挖掘方式。

(1)网络社区知识单元库构建。包括多粒度知识单元识别、知识单元多种关系挖掘、结构化表示与存储3部分,具体内容如下:

多粒度知识单元识别。网络社区中知识单元主要包括3类,即用户、文本与词语,文本又可细分为全文本与句子文本。考虑大多网络社区文本长度较短,传统主题句提取方法作用有限,这里不考虑句子文本,下文所述文本均是指全文本。

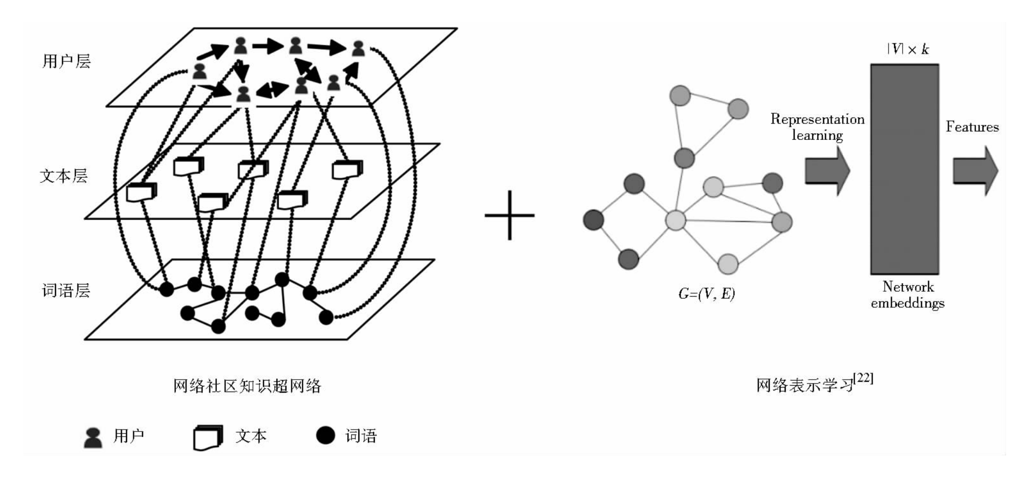


图 1 网络社区多元知识关联体系构建思路

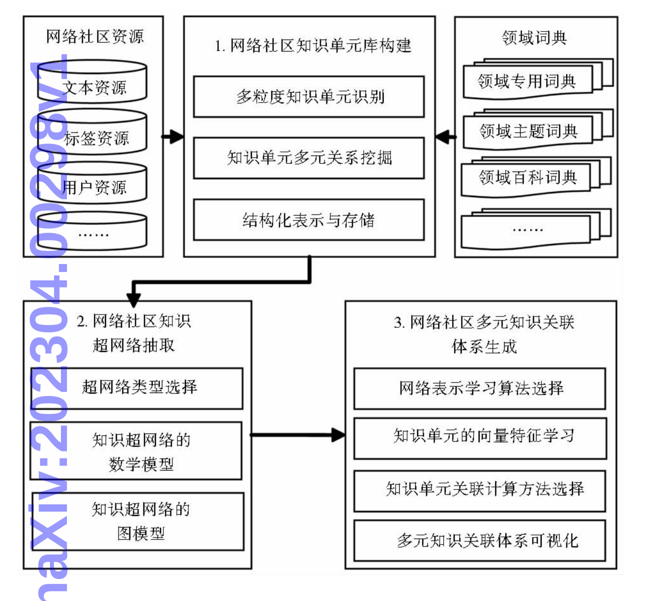


图 2 网络社区多元知识关联体系构建流程

知识单元多元关系挖掘。用户－用户关系分为直接关系与间接关系，本文多元关联体系涉及网络社区多粒度知识单元及其关系，用户之间的间接关系可通过分析用户与词语、用户与文本等关系得到，这里只考虑用户通过回复、关注等行为建立的直接关系；词语－词语关系包括两类：一类是包含用户标签的，可通过统计用户标签共现频次得到，另一类缺乏用户标签的网络资源，可采用主题提取技术（如 LDA 等）或基于领域词典的方式从文本中自动提取；用户－文本关系通过分析用户对文本的操作行为得到，关系类型阈值可人工确定，关系强度由类型阈值与行为强度综合决定；文本－词语关系通过分析词语在文本或文本标签中是否出现得到，关系强度计算方式包括两种，一种只统计词语在文本中出现与否，出现认为有关系，否则无关系，另一种利用文本特征权重（如信息增益、互信息等）进行关系强度细粒度计算。用户－词语关系通过用户－文本关系，文本－词语关系的传递得到。

结构化表示与存储。知识单元存储形式为：

knowledge_unit = <entity, type, description> 式（1）

其中，entity 表示知识单元，type 表示知识单元类型，如文本、用户或词语，description 表示对知识单元的描述。

知识单元关系存储形式为：

knowledge_relationship = <entity1, entity2, relation-type, weight> 式（2）

其中，entity1 与 entity2 分别表示有关系的两个知识单元，relationtype 表示关系类型，weight 表示关系强度。

（2）网络社区知识超网络抽取。网络社区用户、文本与词语知识单元之间存在多种关系类型与关系强度计算方式，仅用一个网络来表征容易导致节点混乱、网络结构不清晰等问题。考虑将其细分成两个节点类型单一、但关系类型异构的网络，即用户关系网络与词语关系网络，之后再利用用户－文本关系与文本－词语关系将两个网络联通，形成节点与关系异构的联通网络。考虑传统异构网络技术在多网络联通表征方面作用有限，笔者选择超网络技术进行网络社区异构知识关系网络构建。美国学者 A. Nagurney 认为超网络是指高于且又超于现存网络的网络[23－24]，一般由多个网络组成，超网络节点可看作网络的集合，边是集合中网络的结合偏好，可通过对边的增加、删除等操作实现对网络结构的调整[12]。

网络社区知识超网络的数学模型与文献[11]中的知识超网络模型类似，只需要将其中句子知识子网络替换成用户知识子网络。知识超网络的图模型如图 3 所示：

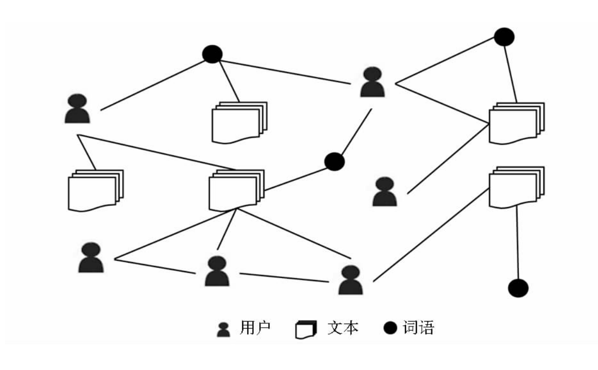


图3　网络社区知识超网络的图模型

（3）网络社区多元知识关联体系生成。网络表示是网络分析的基础，传统表示方法包括基于邻接矩阵与网络图表示两类，前者通过行向量来表征网络节点，容易导致向量维度过高；后者节点间有大量的关联边，使得分析过程倾向采用迭代或组合方式，极大增加算法时间复杂度，最终影响整个分析效果[21]。网络表示学习是一种将网络节点表示到低维空间向量，并利用向量之间距离或相似程度表示节点之间关联的网络表示方式，该方法在极大程度还原初始网络整体结构的同时实现网络的重构，联通了表征现实存在网络与网络分析利用之间的鸿沟，在异构网络的分析利用上优势明显[21-22,25]。笔者利用网络表示学习将网络社区知识超网络中的知识单元表征为低维稠密向量，实现全局视角下的知识单元关联发现，同时降低关系异构对知识组织的不利影响。

网络表示学习算法主要分为基于网络结构与结合外部信息两类，前者包括基于矩阵特征向量计算、矩阵分解、浅层神经网络、深层神经网络等分析算法，后者包括结合文本信息、边的标签信息等分析算法[22]。根据分析对象特点，笔者选择基于浅层神经网络的网络表示学习（LINE）算法，进行网络社区知识单元低维稠密向量表征。LINE 算法解决了 deepwalk 与 node2vec 算法缺乏针对网络结构优化目标函数的问题[13]，同时保留了网络节点的一阶与二阶相似性，对于任意类型的大规模网络都有较高适用性，且由于一阶与二阶相似性的互补，使得该算法能兼顾网络的局部与全部结构[21]。

知识单元关联强度计算是构建多元知识关联体系的关键。在利用 LINE 算法将网络社区知识单元表征成低维稠密向量，关联计算问题就转化为向量相似度计算。当前常用计算方法有余弦相似度、相关系数、欧氏距离、马氏距离等。由于知识单元已表征成低维向量，选择计算绝对距离的欧氏距离来衡量知识单元的关联强度。

4　实证分析

丁香园是国内重要医学社会化媒体平台，论坛用户数与发贴数在学术型网络社区中排名靠前。然而目前论坛资源组织方式仍以传统发布时间、置顶操作等为主，缺乏多维度面向用户与知识内在关联的组织方式。笔者以丁香园论坛中临床医学讨论一区心血管专业讨论版数据为数据源，构建面向心血管领域的多元知识关联体系，为后续知识的深度聚合与组织提供支持。

4.1　数据采集与领域词典构建

检索丁香园心血管论坛[26]的用户发贴信息，检索时间为 2019 年 3 月 17 日。选用火车浏览器抓取用户发贴文本，共得到 65 364 个文本。一个文本保存为一个 TXT 文档，每个文档中包括用户、用户发贴与用户回贴 3 类信息。

由于丁香园心血管论坛尚未提供用户标签功能，笔者利用领域词典识别词语粒度知识单元。综合对比分析考虑以“39 疾病百科”中“心血管内科”栏目[27]为数据源提取心血管领域术语。“39 疾病百科”以信息框方式为每种疾病做了详细的结构化注释，这种结构化注释不但是词典术语的重要补充，也是术语类别划分的重要依据。从网站中共采集到 2 211 个术语，将其划分为疾病（包括病症、别名、并发症）、器官（即发病部位）、症状、诊断（即诊断方法）4 类，具体如表 1 所示：

表1　心血管领域术语词典的数据统计

序号	术语类型	术语个数	类型标签	术语示例
1	疾病	1 177	/njb	冠心病、心绞痛、原发性高血压……
2	器官	93	/nqg	心肌、心脏、血管内皮、心室……
3	症状	652	/nzz	心肌缺血、疼痛、低血钾、紧张……
4	诊断	289	/nzd	血常规、心包积液检查、舒张压……

4.2　丁香园心血管论坛知识单元库构建与知识超网络抽取

笔者从 TXT 文档中分别提取用户与发贴的文本信息，统计共现频次，得到用户-用户关系与关系强度、用户-文本关系与关系强度；然后利用上一步构建的心血管领域术语词典对文本进行分词，通过统计得到文本-词语关系与关系强度、词语-词语关系与关系强度、用户-词语关系与关系强度。具体如表 2 所示：

表 2 丁香园心血管论坛知识超网络的数据统计

知识单元类型	知识单元数(个)	关系类型	累计关系数(次)
用户	101 333	用户 – 用户	7 662 814
文本	65 364	词语 – 词语	10 628 527
词语	1 389	文本 – 用户	379 036
–	–	文本 – 词语	589 747
–	–	用户 – 词语	1 901 896
总计	168 086	总计	21 162 020

4.3 丁香园心血管论坛多元知识关联体系生成

基于上步构建的知识超网络,利用 LINE 算法对丁香园心血管论坛中知识单元进行低维稠密向量表示,向量维度人工设定为 100,得到 168 086 个知识单元的低维稠密向量(部分结果见图 4)。图 4 中第一行表示丁香园心血管论坛多元知识关联体系中共有 168 086 个知识单元,每个知识单元由 100 维向量来表征,除第一行外的第一列为知识单元在关联体系中的 ID 号,其他列为对应向量值。

```
168086 100
2068586 -0.063750 -0.166747 0.096587 -0.038335 -0.090679 0.118372 -0.003603 0.051302 0.041020
2068547 -0.043186 0.073899 0.090871 -0.011568 -0.162485 0.223191 -0.140036 -0.083864 0.107657
2068589 -0.052517 -0.222932 0.112728 -0.183982 -0.037297 0.099639 -0.043205 0.024871 -0.033980
2068581 -0.123077 -0.102651 0.154516 -0.073868 0.054135 0.207710 -0.051798 0.128170 -0.091807
2068490 0.151873 -0.169896 0.229058 -0.158080 0.056862 0.077191 -0.196067 0.094491 0.104568 0.
2068508 -0.111718 -0.148698 0.081895 -0.156985 -0.082217 0.178496 -0.027358 0.079761 0.138894
2068560 0.019455 -0.206083 0.093300 -0.132523 -0.027829 0.141227 -0.054732 0.045444 -0.019019
2068564 -0.061852 -0.082305 0.110416 -0.186133 0.087404 -0.043174 -0.149937 0.103714 -0.076979
2068481 -0.145192 -0.131179 0.175312 -0.231940 -0.004727 -0.012717 -0.157633 0.111545 0.016828
2068430 -0.087248 -0.134144 0.148894 -0.243996 -0.088591 0.064160 -0.156241 0.036792 0.004088
2068592 -0.038122 -0.026459 0.111206 -0.156294 0.012584 0.084853 -0.022706 -0.018120 -0.079673
2068529 -0.110655 -0.080180 0.154456 -0.176822 -0.048632 0.139124 -0.065256 -0.014050 0.015069
2068503 0.003016 -0.150700 0.106419 -0.099971 -0.064676 0.115089 0.008812 0.051971 0.061566 -0
2067542 0.132539 -0.149857 0.184929 -0.130610 0.030856 0.150126 -0.100495 0.033820 0.147701 0.
2068567 -0.126616 -0.095228 0.173882 -0.147645 -0.034810 0.123485 -0.115349 0.056580 0.055289
2068551 -0.065146 -0.168288 0.195792 -0.284619 -0.108883 0.091641 -0.054107 0.007252 0.024515
2068483 -0.074371 0.076664 0.219581 -0.290244 -0.083603 -0.019419 -0.247483 0.302872 0.031501
2067969 -0.159036 -0.049337 0.137277 -0.037975 -0.029338 0.196659 -0.000480 0.131555 -0.171301
2068587 -0.043542 -0.044337 0.096026 -0.162573 -0.074094 0.144297 0.010359 -0.028640 0.046922
2067596 -0.051878 -0.120159 0.146215 -0.231422 -0.182279 0.019115 -0.091597 0.068603 0.037061
```

图 4 丁香园心血管论坛多粒度知识单元的向量表示(部分)

基于知识单元的特征向量,选用欧氏距离计算 168 086 个知识单元的关联强度,得到丁香园心血管论坛的多元知识关联体系。以疾病“高血压”为例展示词语知识单元在多元关联体系中的多元关联集合,具体如表 3 所示。值得一提的是,笔者在构建术语词典时对术语进行了类型标注,因此多元关联体系中的词语关联除了包含关联强度外还包含关联类型,这种细粒度关联类型是支持网络社区高级知识服务基础。

表 3 “高血压”的多元知识关联集合(部分)

排序	关联文本 ID	欧氏距离	关联用户	欧氏距离	关联词语	欧氏距离	关联类型
1	5 204 018	1.373 744 05	小人物人小	1.013 477 049	糖尿病	0.451 070 467	疾病 – 疾病
2	20 857 874	1.378 115 799	zhangjunbo1973	1.119 331 951	血糖	0.929 704 086	疾病 – 诊断
3	12 640 619	1.379 826 709	heaven197898	1.125 586 955	紧张	0.930 057 62	疾病 – 症状
4	4 744 009	1.380 998 061	青柳御前	1.134 284 495	冠心病	0.930 181 917	疾病 – 疾病
5	5 484 572	1.381 039 39	wangyy1990	1.136 411 634	高血压病	0.994 399 514	疾病 – 疾病
6	26 449 216	1.388 507 339	竹枝 9423	1.137 009 612	血管	1.045 712 232	疾病 – 器官
7	11 429 148	1.389 119 356	xjxianjun	1.144 390 778	肥胖	1.067 830 842	疾病 – 症状
8	18 509 594	1.389 776 15	desperado – c	1.145 501 249	综合征	1.081 221 046	疾病 – 疾病
9	2 550 115	1.390 046 058	diasy	1.146 014 646	高血脂	1.085 656 541	疾病 – 疾病
10	16 018 737	1.390 173 125	ahmatdr	1.159 744 804	低血压	1.094 429 361	疾病 – 疾病

为了更好可视化表示多元知识关联体系,利用 PCA(主成分分析)进行降维处理,然后利用 Python 中的 Matplotlib 进行可视化展示。以“高血压”为例,可视化展示词语知识单元的多元知识关联集合,具体见图 5。

除此之外,为了更好地比较多元知识关联体系与知识超网络的优劣,以上一节抽取的知识超网络为基础衡量知识单元的关联度。目前,网络节点相似度计算方法包括基于网络拓扑结构[28]、节点属性以及两者的综合[29]。为简化计算过程考虑采用基于节点属性方法,首先构建知识单元的属性向量,然后选用皮尔逊系数衡量知识单元关联度,得到基于超网络的知识单

肖璐，赵之辉，陈果. 全局视角下的网络社区多元知识关联挖掘[J]. 图书情报工作,2020,64(6):100－107.

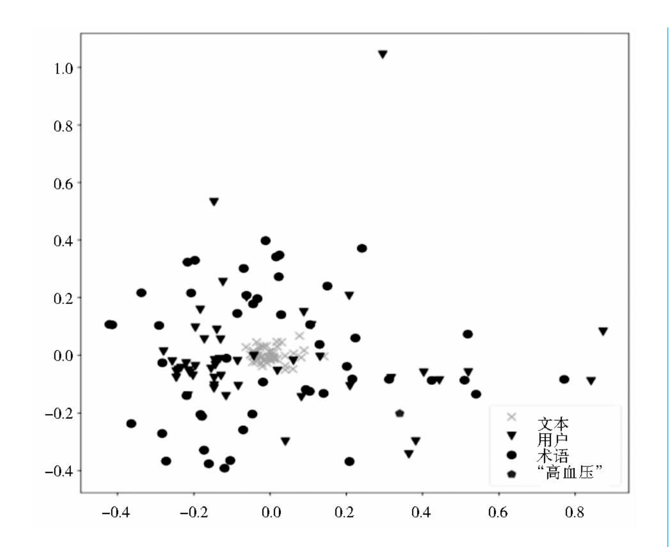


图 5 “高血压”的多元知识关联集合可视化结果

元关联集合。为清晰对比挖掘效果，这里只列出与“高血压”关联度高的 10 个词语，具体如表 4 所示。对比表 3 与表 4 可知，基于表示学习方法挖掘的关联词语类型更全，包括 4 类术语；且挖掘出了与“高血压”具有潜在关联的术语“糖尿病”“冠心病”等。糖尿病与高血压是同源性疾病[30]，目前已有很多研究针对冠心病与高血压合并情况开展[31－32]；除此之外，表 4 中“颈椎 CT 检查”一词，经领域专家判断，认为与“高血压”关联不高。由此可知本文多元知识关联挖掘方法的有效性。

表 4 基于知识超网络的“高血压”高关联词语集合（部分）

排序	关联词语	皮尔逊系数	关联类型
1	高血脂	0.404 03	疾病－疾病
2	颈椎 CT 检查	0.398 32	疾病－诊断
3	高血压肾病	0.396 94	疾病－疾病
4	血流阻力增加	0.393 84	疾病－症状
5	血压波动大	0.391 19	疾病－症状
6	眼底检查	0.375 51	疾病－诊断
7	肾图检查	0.374 36	疾病－诊断
8	血液粘度增高	0.371 32	疾病－症状
9	老年人高血压	0.370 16	疾病－疾病
10	血压高	0.369 69	疾病－症状

4.4 丁香园心血管论坛多维知识聚合原型系统设计

目前，丁香园心血管论坛资源组织以置顶操作等为主，且未提供针对分论坛的检索功能。以“高血压”为关键词进行全社区资源检索，从结果中提取心血管论坛数据，得到图 6 所示结果。这种基于关键词匹配的列表式资源检索模式难以满足用户高级知识服务需求，因此考虑以多元知识关联体系为基础，挖掘资源内在关联，通过对命中资源的多维聚合组织实现多维知识服务。以关键词“高血压”的检索结果为例，展示多维聚合结果，该结果通过挖掘丁香园心血管论坛多元知识关联体系并人工调整得到，具体如图 7 所示。图 7 左边是一级检索结果，分别从文档、术语与用户 3 个维度展示了与“高血压”高关联资源；右边是检索结果的二级扩展，其中①为文档“2010 年高血压防治指南”的文档维度扩展；②为“高血压心电图疑惑：为何出现顺种向转位”文档中“心电图”的术语维度扩展；③为术语维度检索结果“冠心病”的二级扩展；④为一级检索结果用户“heaven197898”的文档维度扩展。通过多维知识聚合用户除了可以获取目标文档外还可以进行相关术语扩展与用户识别，实现论坛的高级知识服务。

《高血压图谱（第6版）》
《高血压图谱(第6版)》是临床医师，特别是研究治疗高血压病的专业医师和研究人员很有价值的参考书。作者简介作者:(美)郝伦伯格 著作 高玖鸣 译者:高玖鸣目录第1章高血压的发病机制：遗传与环境因素1
丁香园 - 心血管 - 2018-10-29 13:11:51 38

《顽固性高血压》pdf，高血压学科丛书，余振球
本书从各个角度对顽固性高血压进行了阐述，主要介绍顽固性高血压的 界定、假性顽固性高血压的重要性，尤其对顽固性高血压的原因做了全面、 系统、详细的描述。鉴于继发性高血压在顽固性高血压中所占比例较大
丁香园 - 心血管 - 2017-04-28 12:56:51 42

高血压各种情况下的选药策略（完）
CCB对心衰患者没有有益的证据。如必须使用二氢吡啶类CCB，可用氨氯地平或非洛地平。高血压所致的心衰以舒张功能不全为主，大剂量的洋地黄可导致心肌浆网中钙离子超负荷，反而损害心肌，降低心肌顺应性，加重
丁香园 - 心血管 - 2018-08-02 14:28:56 153

图 6 丁香园心血管论坛“高血压”检索结果（部分）

5 结语

针对网络社区中多元知识关联挖掘面临的难题，笔者提出在超网络的基础上开展网络表示学习的方案。以丁香园心血管论坛为例开展的实验表明，该方案有如下优点：(1) 知识单元间的关联挖掘全面参照了其与用户、领域术语、文本的关系，而非单一种类关系，这种全局视角下的知识关联挖掘结果更为可靠；(2) 将用户、领域术语、文本等不同类型知识单元转化为同一特征空间下的低维稠密向量，故而在后续知识关联的计算中可屏蔽知识单元类型差异、关系异质的干扰，多元关联的挖掘简洁有效；(3) 保留了知识单元类型，因此所得的知识关联除强度外，仍保留了类型差异（如“用户－术语”“用户－用户”等），后续知识组织中可根据应用场景有效区分，例如，给定一个用户，根据其 top N 个最相关术语对其打标签，根据 top N 个最相关用户对其开展用户推荐，根据 top N 个最相关帖子对其开展资源推荐。

本文研究不足之处是在超网络构建和后续网络表示学习中，没有考虑网络节点的文本内容，目前已有研

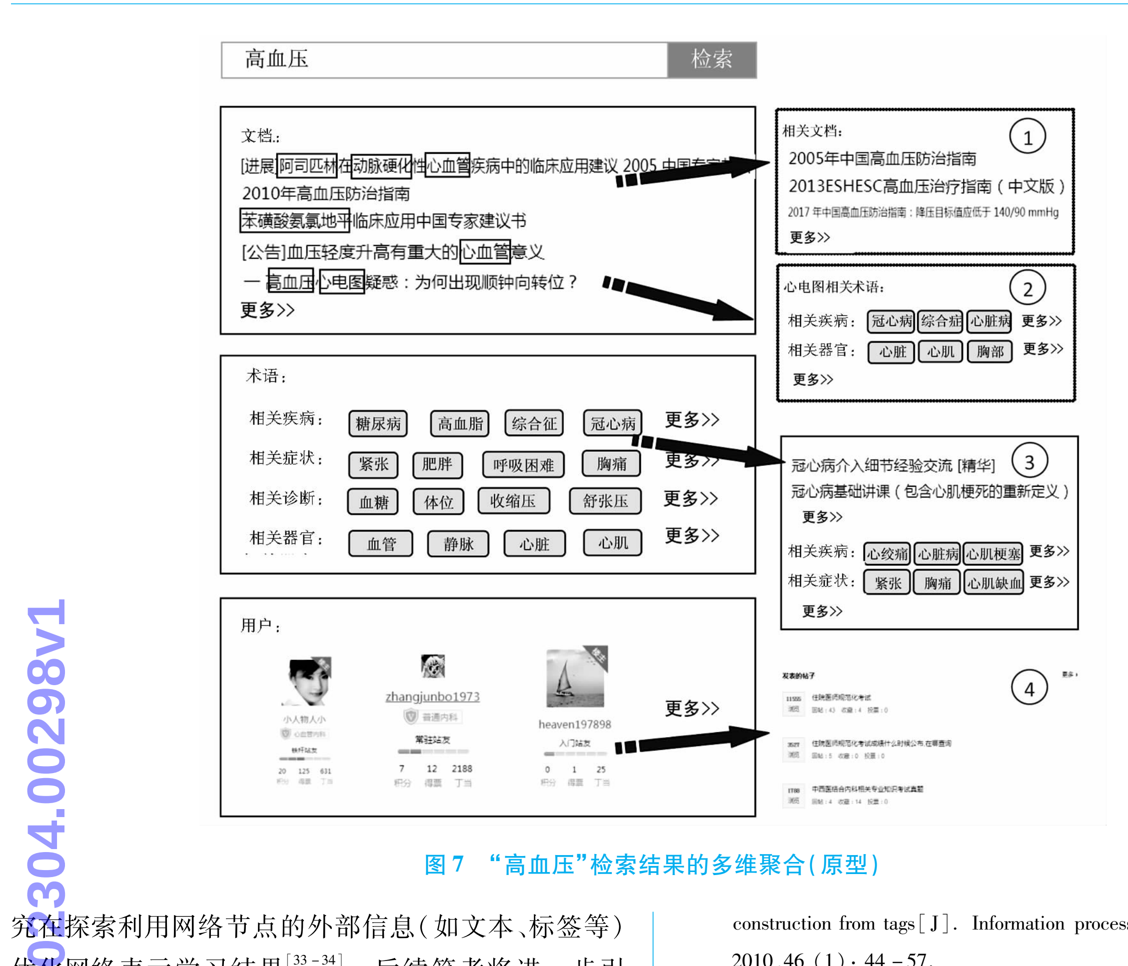


图 7　“高血压”检索结果的多维聚合(原型)

究在探索利用网络节点的外部信息(如文本、标签等)优化网络表示学习结果[33-34]。后续笔者将进一步引入结合外部信息的网络表示学习方法开展网络社区多元知识关联挖掘。

参考文献:

[1] 陈果,朱茜凌,肖璐. 面向网络社区的知识聚合:发展、研究基础与展望[J]. 情报杂志,2017,36(12):193-197,192.

[2] 田博,凡玲玲. 基于交互行为的在线社会网络社区发现方法研究[J]. 情报杂志,2016,35(11):183-188.

[3] 刘冰玉,王翠荣,王聪,等. 基于动态主题模型融合多维数据的微博社区发现算法[J]. 软件学报,2017,28(2):246-261.

[4] 宁亚辉,樊兴华,吴渝. 基于领域词语本体的短文本分类[J]. 计算机科学,2009,36(3):142-145.

[5] SAHAMI M, HEILMAN T D. Aweb-based kernel function for measuring thesimilarity of short textsnippets[C]//Proceedings of the 15thinternational conference on World Wide Web. New York: ACM, 2006:377-386.

[6] ZELIKOVITZ S, KOGAN M. Using web searches on important words to create background sets for LSI classification[C]//Proceedings of the 19th international FLAIRS conference. Florida: AAAI Press,2006: 598-603.

[7] 黄微,高俊峰,李瑞,等. Folksonomy 中 Tag 语义距离测度与可视化研究[J]. 现代图书情报技术,2014(7/8):64-70.

[8] TSUI E, WANG W M, CHEUNG C F, et al. A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags[J]. Information processing &management, 2010,46 (1): 44-57.

[9] 陈果. 基于领域概念关联的网络社区知识聚合研究[D]. 武汉:武汉大学,2015.

[10] 石伟杰,徐雅斌. 微博用户兴趣发现研究[J]. 现代图书情报技术,2015(1):52-58.

[11] 肖璐. 基于知识超网络的网络社区学术资源多粒度聚合研究[J]. 情报杂志,2018,37(12):182-187,194.

[12] 王传清,毕强. 超网络视域下的数字资源深度聚合研究[J]. 情报学报,2015,34(1):4-13.

[13] 张金柱,于文倩,刘菁婕,等. 基于网络表示学习的科研合作预测研究[J]. 情报学报,2018,37(2):132-139.

[14] 林原,刘海峰,王海龙,等. 基于表示学习的学者间潜在合作机会挖掘[J]. 情报杂志,2019,38(5):65-70.

[15] 樊玮,韩佳宁,张宇翔. 基于网络表示学习的论文影响力预测算法[J]. 计算机工程,2019,45(12):160-165,170.

[16] 张潇鲲,刘琰,陈静. 引入外部词向量的文本信息网络表示学习[J]. 智能系统学报,2019,14(5):1056-1063.

[17] 朱国进,李承前. 网络知识资源表示学习模型[J]. 智能计算机与应用,2016,6(3):5-10.

[18] 朱靖雯,杨玉基,许斌,等. 基于 HowNet 的语义表示学习[J]. 中文信息学报,2019,33(3):33-41.

[19] 韩忠明,郑晨烨,段大高,等. 基于多信息融合表示学习的关联用户挖掘算法[J]. 计算机科学,2019,46(4):77-82.

[20] 杨奕卓,于洪涛,黄瑞阳,等. 基于融合表示学习的跨社交网络用户身份匹配[J]. 计算机工程,2018,44(9):45-51.

[21] 尹赢,吉立新,黄瑞阳,等.网络表示学习的研究与发展[J].网络与信息安全学报,2019,5(2):77－87.
[22] 涂存超,杨成,刘知远,等.网络表示学习综述[J].中国科学:信息科学,2017,47(8):980－996.
[23] NAGURNEY A,DONG J. Supernetworks: decision-making for the information age[M]. Cheltenham: Edward Elgar Publishers, 2002.
[24] 漆玉虎,郭进利.超网络研究[J].上海理工大学学报,2013,35(3):227－239.
[25] 周慧,赵中英,李超.面向异质信息网络的表示学习方法研究综述[J].计算机科学与探索,2019,13(7):1081－1093.
[26] 丁香园论坛.心血管专业讨论版[EB/OL].[2019－06－08].http://cardiovascular.dxy.cn/bbs/board/47.
[27] 39疾病百科.心血管内科[EB/OL].[2019－06－08].http://jbk.39.net/bw/xinxueguanneike.
[28] 张良富,李翠平,陈红.大规模图上的SimRank计算研究综述[J].计算机学报,2019,42(12):2665－2682.
[29] 邱少明,於涛,杜秀丽,等.基于节点多属性相似凝聚的社团划分算法[J/OL].[2019－10－13].http://kns.cnki.net/kcms/detail/31.1289.tp.20190808.1701.023.html.
[30] 百度百科.糖尿病高血压[EB/OL].[2019－10－13].https://baike.baidu.com/item/%E7%B3%96%E5%B0%BF%E7%97%85%E9%AB%98%E8%A1%80%E5%8E%8B/1424758?fr=aladdin.
[31] 张菀桐,胡元会,朱宝琛,等.冠心病合并高血压患者血压水平与血栓形成动力学相关性研究[J].现代中西医结合杂志,2016,25(2):129－131,140.
[32] 张鑫,李荣,黄玉晓,等.OSAHS对冠心病合并高血压患者动态血压及心率变异性的影响[J].中西医结合心脑血管病杂志,2014,12(1):35－37.
[33] YANG C,LIU Z Y,ZHAO D, et al. Network representation learning with rich text information[C]//Proceedings of the 24th international joint conference on arti? cial intelligence. Buenos Aires: AAAI Press, 2015:2111－2117.
[34] TU C C,LIU H,LIU Z Y,et al. CANE: context-aware network embedding for relation modeling[C]//Proceedings ofthe 55th annual meeting of the association for computational linguistics. Vancouver: ACL, 2017:1722－1731.

作者贡献说明：
肖璐：撰写与修改论文、论文定稿；
赵之辉：参与论文讨论、整理文献；
陈果：修改论文、提供领域词典。

Holistic Perspective Multi-knowledge Relations Mining in Network Community

Xiao Lu[1] Zhao Zhihui[1] Chen Guo[2]
[1] School of Journalism, Nanjing University of Finance & Economics, Nanjing 210023
[2] School of Economics & Management, Nanjing University of Science and Technology, Nanjing 210094

Abstract: [**Purpose/significance**] There are many knowledge units in the network community, among which there are intricate relationships. It is necessary to carry out multiple knowledge relations mining uniformly and succinctly on the premise of retaining all the relations of knowledge units. [**Method/process**] This paper puts forward the solution of multi-knowledge relations mining in network community. Firstly, 3 typical knowledge units (users, texts and words) in the network community and their multiple relations in the knowledge communication were extracted into a supernetwork. Secondly, the network representation learning algorithm was used to uniformly represent the nodes in the supernetwork as low-dimensional dense vectors. Finally, multiple knowledge relations calculation was carried out based on nodal vector. [**Result/conclusion**] The effectiveness of the scheme was verified by taking cardiovascular BBS in dingxiang garden as an example. This scheme not only retains all the information of the knowledge unit, but also carries out the mining of the knowledge relation under the unified low-dimensional characteristics, and finally the knowledge relation meets the requirements of the diversity of the knowledge organization scene in the network community.

Keywords: knowledge relation mining　super network　network representation learning　network community